非ISA关系在本体概念相似度计算中的度量方法研究

王孝满 闫晶晶
(信息系统工程重点实验室 南京 210007)

摘 要 在本体概念相似度计算过程中,关于本体概念间非ISA关系的处理方法较少。针对本体中存在非ISA概念关系的情况,总结了一些传统的概念相似度计算方法,提出了一种新的适应于非ISA概念关系的相似度计算方法。此方法利用Tversky模型计算本体有向无环概念图的信息量覆盖程度,并结合语义距离方法,再进行权值求和。实验表明,提出的方法可以有效地度量ISA关系,对非ISA关系具有适用性。

关键词 非ISA概念关系,本体概念相似度,信息量,概念图
中图法分类号 TP301 **文献标识码** A

Approach of NON-ISA Concept Relation in Ontology Concept Similarity Computation

WANG Xiao-man YAN Jing-jing
(Science and Technology on Information Systems Engineering Laboratory, Nanjing 210007, China)

Abstract In the process of ontology concept similarity computation, there is few about the approaches of dealing with non-isa concept relation. Considering the existence of non-isa concept relation in ontology, summarize some traditional approaches of concept similarity computation, proposed a novel approach which adapts the non-isa relation. Compute the overlay-grade of ontology directed acycline concept graph information content based on tversky model, and combined the semantic distance approach, then sumed the weighted value. Experimental results show that our approach is effectively for isa concept relation and applicable for non-isa concept relation.

Keywords NON-ISA concept relation, Ontology concept similarity, Information content, Concept graph

1 引言

相似度计算被广泛应用于语义检索[1]、本体映射[2]、词义消歧[3,4]、语义Web服务匹配[5]等领域,是很多相关应用的前提。在相似度计算中,本体概念相似度计算是本体相关应用的关键,其传统的方法很少有详细讨论非ISA概念关系对相似度计算的影响,如part-of,value-of等。有些传统方法在存在非ISA关系的本体概念相似度计算中失效,制约了相关的应用。为有效地度量非ISA概念关系,有必要提出一种新的度量方法。本文总结了传统的相似度计算方法,提出利用Tversky[6]模型的概念图信息量覆盖度来衡量本体概念相似度,同时结合概念间的语义距离信息,综合地计算存在非ISA关系的本体概念相似度。实验表明,本文方法一方面在存在ISA关系的本体概念相似度的计算中可取得较好效果,另一方面适用于非ISA概念关系,可以克服一些传统方法赋予非ISA关系固定权值的缺陷。

2 传统的相似度计算方法

传统的相似度计算可以分为基于知识的方法、基于文本的方法、gloss overlap方法。

2.1 基于知识的方法

基于知识的方法又可以分为基于边的方法、基于节点的方法。

基于边的方法,也即基于路径的方法,在仅存ISA关系的本体中,最先计算概念之间的边的数目来计算相似度[7],接着引入概念最近共同祖先节点的深度[8],Hirst[9]认为距离越近且其路径中方向变化越少,概念越相似。文献[10]综合考虑路径长度、深度、密度等对相似度的影响。文献[11]利用概念距离定义了惩罚因子,改进了文献[7]的方法。

为了克服基于边的方法的不可靠性,出现了基于节点的信息来计算相似度的方法。节点信息可以是节点特征描述信息或信息量,据此基于节点的方法分为基于特征的方法和基于信息量的方法。

2.1.1 基于特征的方法

基于特征的方法,即基于Tversky模型,或者是其变型[12],如Jaccard,Dice,inclusion等。Tversky模型是

$$\mathrm{sim}_{tver}(A,B)=\theta f(A\cap B)-\alpha f(A-B)-\beta f(B-A) \quad (1)$$

2.1.2 基于信息量的方法

基于信息量的方法首先由Resnik[13]提出。在WordNet[14]名词的上下位关系中,概念越抽象,信息量越小,概念越具体,信息量越大。信息量的量化值是通过概念在一个文档集中出现的概率来计算的

$$IC(c)=-\log(p(c)) \quad (2)$$

到稿日期:2010-08-23 返修日期:2010-12-10 本文受国家重点实验室创新基金课题(9140C8301011001)资助。
王孝满(1986-),男,硕士生,主要研究方向为语义网、本体、语义Web服务,E-mail:wxm5558@163.com;闫晶晶(1977-),女,高工,主要研究方向为软件总体和辅助决策技术。

式中，$p(c)$是指概念 c 在特定训练文档集中出现的概率。Resnik 认为两个概念共享的信息量越多，其越相似。基于信息量的相似度计算公式为

$$\mathrm{sim}_{res}(c_1,c_2)=\max_{c\in S(c_1,c_2)}[-\log(p(c))] \tag{3}$$

式中，$S(c_1,c_2)$是两个概念的所有祖先概念集合。这种方法不能区分两对概念具有相同父概念集的情形。Jiang & Conrath[15]对其进行了改进，对每个 ISA 关系连接赋予一个连接强度 LS(link strength)，并提出了语义距离

$$Dist_{jcn}(w_1,w_2)=IC(c_1)+IC(c_2)-2\times IC(LSuper(c_1,c_2)) \tag{4}$$

D. Lin[16]提出了类似的方法，考虑共同最近祖先的信息量以及概念自身的信息量

$$\mathrm{sim}_{lin}(c_1,c_2)=\frac{2\times IC(lcs(c_1,c_2))}{IC(c_1)+IC(c_2)} \tag{5}$$

式中，$lcs(c_1,c_2)$是概念 c_1,c_2 的最近共同祖先。Nuno 等[17]提出，概念的下位关系越多，其信息量就越小，叶子节点应该包含最大的信息量，给出了 WordNet 中的节点信息量的计算公式为

$$IC_{wn}(c)=\frac{\log\left(\frac{hypo(c)+1}{\max_{wn}}\right)}{\log\left(\frac{1}{\max_{wn}}\right)}=1-\frac{\log(hypo(c)+1)}{\log(\max_{wn})} \tag{6}$$

式中，$\max_{wn}$是词典中含有的词汇总数量，$hypo(c)$表示概念 c 的下位概念数。文献[18]提出了与文献[17]类似的信息量度量方式，结合 Tversky 模型提出概念相似度计算方法

$$\mathrm{sim}_{tw'}(c_1,c_2)=3\times IC(msca(c_1,c_2))-IC(c_1)-IC(c_2) \tag{7}$$

式中，$msca(c_1,c_2)$是最近共同父概念。以上信息量的计算是基于固定文本集。S. David[19]提出了基于 Web 网页计数的信息量定义方式

$$IC\_IR(a)=-\log p_{web}(a)=-\log\frac{hits(a)}{total\_webs} \tag{8}$$

式中，$p_{web}(a)$指 a 在 Web 搜索结果中出现的概率，$total\_webs$ 是 Web 搜索引擎索引的资源总数。基于这个信息量定义方式，重新改写了 Resnik，Lin，Jiang&Conrath 的相似度计算公式。基于 Web 的信息量定义方式可以克服基于文本集的一些缺点，如特定领域的词汇不能覆盖。

2.1.3 gloss overlap 方法

使用 WordNet 作为知识源，查询概念的词汇在 WordNet 中对应的词义项的 gloss，比较 gloss 文本的相关度作为概念之间的相似度，例如 Lesk[20]算法、extended gloss overlap[21]方法以及文献[2]提出的几个 gloss overlap 的匹配器。

2.2 基于文本的方法

基于文本的方法是指通过自然语言处理技术，分析大量文档的统计规律来计算相似度的方法。可以分为 5 类：基于概念词汇共现率的计算方法，如互信息方法[22]；基于上下文的计算方法，如 context dependency 方法[23]；基于 Web 的方法，如基于 Web 搜索引擎的计算方法[24,25]；基于维基百科知识库[26]的方法；潜语义分析 LSA。限于篇幅，不再详细讨论，只重点讨论与本文有关的 2.1 节的部分。

3 相关工作

文献[27]将本体概念相似度分为概念相似度和描述相似度，而描述相似度包括关系相似度和属性相似度。其中关系相似度并没有严格区分 ISA 与非 ISA 关系，对于特定的关系赋予一个特定的权值。文献[28]指出两个节点间的边的权重为

$$w(A,B)=\frac{w(A\rightarrow_r B)+w(B\rightarrow_{r'}A)}{2d}$$

式中，$w(X\rightarrow_r Y)=\max_r-\frac{\max_r-\min_r}{n_r(X)}\rightarrow_r$ 是边的类型，$\max_r$，$\min_r$ 分别为关系 r 的最大与最小权重，$n_r(X)$表示节点 X 的出度。文献[29]对边赋予一个固定的权重

$$weight(a,b)\propto\begin{cases}1, & type(a,b)=synonymy\\ 0.95, & type(a,b)=hyper-hyponym\\ 0.9, & type(a,b)=hol-meronym\end{cases} \tag{9}$$

文献[30]中考虑其他关系如 part-of，Associative 等，但仍是赋予一个固定的权重。文献[31]将语义路径分为 SR(single-relation path)和 MR(mixed-relation path)，对于非 ISA 关系也是赋予一个固定的权重，对特定关系赋予特定的权重，不具有普遍性。

文献[32]利用本体中概念和关系相互增强的迭代方式计算概念重要性和关系权重，提出了相互迭代的算法，求得关系的重要性权重，但其仅仅是从概念之间的连接结构特性进行关系重要性判断，与概念自身的特性分离了。文献[33]指出 WordNet 中存在 9 种类型的边，以信息量为基础，讨论了上下文关系(ISA)的边权重计算特性，边的平均连接强度为

$$avg(arc)=\frac{\sum_{i=1}^{n}\{IC(c_i)\times(oc_{hyper}-oc_{hypon})\}}{n-1} \tag{10}$$

式中，oc_{hyper}为 c_i 上位概念数，oc_{hypon}为 c_i 下位概念数，n 为节点总数。但其没有给出其他的非 ISA 关系的边连接强度的计算方法。文献[34]提出 common specific(CSpec)概念，结合信息量和 CSpec 共同给出语义距离衡量方式，但是基于 ISA 关系，没有考虑非 ISA 关系。文献[35]将概念之间的语义距离定义为蕴含距离和定义距离，其前提是在 ISA 关系层次中。文献[36]也是以 ISA 关系为前提，通过匹配概率大小进行精确的相似度传播，该方法对于一些非 ISA 关系不能适应。

综合上述各种研究，传统的本体概念相似度计算中存在两个问题：

(1)部分方法没有考虑非 ISA 关系的影响，对于存在非 ISA 关系的本体不适用。

(2)有些方法考虑非 ISA 关系，但只是简单地对特定关系的边赋予一个特定的权重，不具有普遍性。

本文提出的本体概念相似度度量方法，利用 Web 的网页计数衡量概念信息量，相对于传统的基于特定文本集的方式具有公正性和独立性，对 ISA 和非 ISA 都适用，概念词汇覆盖率更高，克服了赋权重的缺陷，具有普遍性。

4 非 ISA 关系的本体概念相似度度量

4.1 相关定义

定义 1(本体有向无环图) 本体可以表示为有向无环图 $G=\langle V,E\rangle$，其中 V 是本体中全体概念节点的集合，E 是概念之间的关系的集合，即连接概念节点的有向边。

定义 2(路径) 本体有向无环图 $G=\langle V,E\rangle$，对于任意两个节点 $v_i\in V,v_j\in V$ 之间的连通有向路径 $p=\langle v_i,v_{i+1},\cdots,$

$v_j\rangle$，其中 v_i 是 v_{i+1} 的直接父概念，路径 p 是连接 v_i 与 v_j 的有向边的集合，p 可以存在多个。

定义 3（入度与出度） 在有向无环图 G 中，对于节点 $v\in V$，以 v 为头的边的数目称为 v 的入度，记为 $ID(v)$；以 v 为尾的边的数目称为 v 的出度，记为 $OD(v)$。根据有向无环图的性质，可以得到

$$\sum_{v_i\in V} ID(v_i)=\sum_{v_i\in V} OD(v_i)=|E| \tag{11}$$

即所有节点的入度之和、出度之和都与边的数量相等。

文献[32]指出在本体的有向无环图中，一个概念节点作用于其它概念的边越多，则该概念越重要，即概念出度越大，则概念越重要；入度越大，概念也越重要。

定义 4（重要性） 在本体有向无环图 $G=\langle V,E\rangle$ 中，对于任意概念节点 $v_i\in V$，v_i 的重要性表示为 r_i，$\sum_{i=1}^{|V|} r_i=1$，其中 $1\leqslant i\leqslant|V|$。概念节点重要性的计算公式为

$$r_i=(ID(v_i)+OD(v_i))/2\times|E| \tag{12}$$

4.2 新的相似度度量方法

许多传统的本体概念相似度计算主要是针对只包含 ISA 关系的本体，综合考虑本体概念的深度、密度、祖先节点等因素。在包含非 ISA 关系的本体有向无环图中，概念的深度、密度因素对概念相似度的影响是不确定的。因此，本文仅考虑节点信息量、语义距离两个因素，采用基于语义距离和概念图信息量覆盖率的方法，不限制于本体概念关系类型。

4.2.1 概念图覆盖

根据本体有向无环图性质，提出基于 Tversky 模型的概念图覆盖率的相似度计算方法，首先给出相关定义。

定义 5（概念图） 对于任意节点 $v'\in V$ 的概念图 $G'=\langle V',E'\rangle$，其中 V' 表示至少和 v' 有一条路径的节点的集合，E' 表示只连接 V' 中节点的边的集合。

对于本体有向无环图中的任意两个概念 v_i,v_j 之间的相似度，可以表示为

$$\mathrm{sim}_{cg}(v_i,v_j)=\frac{\alpha f(G_{v_i}\cap G_{v_j})}{\alpha f(G_{v_i}\cap G_{v_j})+\beta f(G_{v_i}/G_{v_j})+\gamma f(G_{v_j}/G_{v_i})} \tag{13}$$

式中，$\mathrm{sim}_{cg}(v_i,v_j)$ 表示概念 v_i,v_j 的概念图覆盖率相似度；G_{v_i},G_{v_j} 分别表示概念 v_i,v_j 的概念图；$G_{v_i}\cap G_{v_j}$ 表示两个概念图的交集；G_{v_i}/G_{v_j} 表示属于概念图 G_{v_i} 不属于 G_{v_j} 的部分；G_{v_j}/G_{v_i} 表示属于 G_{v_j} 不属于 G_{v_i} 的部分；函数 $f(G)$ 表示概念图的所有概念节点的信息量之和；α,β,γ 是调节参数。

节点的重要性越高，则与其相关的边就越重要。在相似度计算中，重要节点的信息量应起到主要作用，因此 3 个调节参数分别是对应节点的重要性比例，表示为

$$\begin{cases}\alpha=\dfrac{\sum_{v_i\in(G_a\cap G_b)} r_i}{M}\\ \beta=\dfrac{\sum_{v_i\in(G_a/G_b)} r_i}{M}\\ \gamma=\dfrac{\sum_{v_i\in(G_b/G_a)} r_i}{M}\end{cases},\ M=\sum_{v_i\in(G_a\cup G_b)} r_i \tag{14}$$

4.2.2 语义距离

概念之间的相似度可以由语义距离来反映。传统的语义距离是通过边的连接强度之和来衡量的，两个概念在本体中的路径越短，就越相似。每条边的连接强度由概念的信息量之差决定[15]。基于语料库术语出现概率的信息量计算方法，计算结果与语料库的特性相关[13]。根据本体下位概念数量计算内在的信息量[17]，前提是概念之间是 ISA 关系。因此，为了适应非 ISA 关系的影响，本文采用基于 Web 网页计数的信息量计算方式[19]。

定义 6（边的强度） 边的强度即是连接两个概念的边的距离权重，可以通过概念的信息量之差来衡量。给定两个相邻概念节点 v_i,v_j，采用式(8)的信息量计算方法，边的强度为

$$es(v_i,v_j)=|IC\_IR(v_i)-IC\_IR(v_j)| \tag{15}$$

在非 ISA 关系中，子概念与父概念信息量大小不是确定的，边的强度为信息量差的绝对值。

定义 7（语义距离） 语义距离是路径中边的强度之和。由于概念之间可能存在多条路径，对于一个特定的路径 P_x，$parent(v')$ 表示 v' 的直接父概念，语义距离可以表示为

$$dist_{P_x}(v_i,v_j)=\sum_{v'\in P_x(v_i,v_j)} es(v',parent(v')) \tag{16}$$

取最小的语义距离 $dist(v_i,v_j)=\min_{P_x\in P}(dist_{P_x}(v_i,v_j))$，$P$ 表示概念节点 v_i,v_j 的所有路径集合。结合概念图相似度和语义距离，得到本体概念相似度的度量方式，$0<\lambda<1$ 为调节因子。

$$\mathrm{sim}(v_i,v_j)=\lambda\times\mathrm{sim}_{cg}(v_i,v_j)+\frac{1-\lambda}{(dist(v_i,v_j)+1)} \tag{17}$$

4.3 一个实例

手工构建了一个关于车的本体，如图 1 所示。其中本体概念之间的关系除了 ISA 关系，还包括 CHR(characterized by)，part-of，LOC(location)。采用本文方法，计算节点 *ferrari* 与 *safe* 的相似度。*ferrari* 的概念图 $G_{ferrai}=\{thing,vehicle,car,station\ wagon,ferrari\}$，*safe* 的概念图 $G_{safe}=\{thing,vehicle,car,SUV,volvo,safe\}$，根据式(14)可得 $\alpha=14/25,\beta=5/25,\gamma=6/25$。

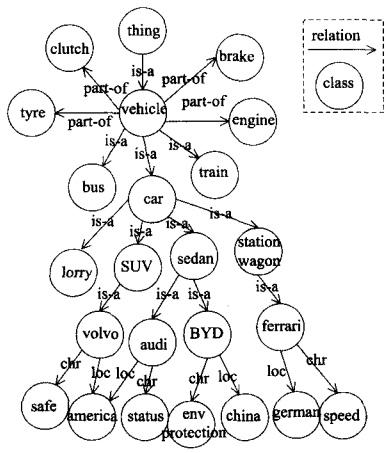

图 1 关于车的本体有向无环图

表 1 节点网页计数与信息量

节点	网页计数	信息量
Thing	1250000000	0.9031
Vehicle	376000000	1.4248
Car	1560000000	0.8069
Station wagon	5480000	3.2612
Ferrari	114000000	1.9431
SUV	70900000	2.1494
Volvo	125000000	1.9031
Safe	680000000	1.1675
Speed	799000000	1.0975

利用 Web 搜索引擎获取概念词汇的网页计数。本文使用 Google[37]，假设 Google 全部索引网页计数为 100 亿，同时计算出节点信息量，如表 1 所列。

可以计算出基于概念图覆盖的概念相似度为

$$sim_{cg}(ferrari,safe)=\frac{\alpha f(G_{ferrari}\cap G_{safe})}{\alpha f(G_{ferrari}\cap G_{safe})+\beta f(G_{ferrari}/G_{safe})+\gamma f(G_{safe}/G_{ferrari})}=0.7742$$

计算出语义距离为 $dist(ferrari,safe)=6.0968$，取 $\lambda=0.5$，代入式(18)，得 $sim(ferrari,safe)=0.4576$，$sim(ferrari,speed)=0.7675$。由此可见，人们对于法拉利安全性的关注较少，对于速度关注多，符合实际思维习惯。如果采用传统的方法[13,15-18]，不能有效计算这两对概念之间的相似度，因为它们不能处理 loc，chr 等非 ISA 概念关系。

5 实验

采用 Washington 大学的 FMA[38] (Foundational Model of Anatomy)项目的本体。FMA 本体概念之间的关系以 ISA 和 part-of 关系为主，还有很多其他关系，如 has_shape，has_boundary，connected_to 等。手工选择其中的部分概念术语，构建了实验本体，如图 2 所示。

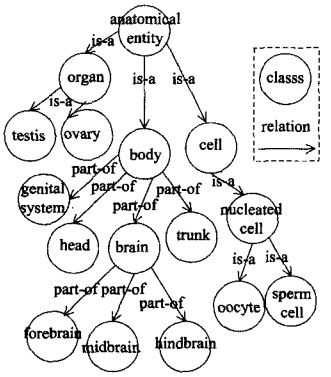

图 2 FMA 部分术语构成的本体

表 2 ISA 关系的相似度比较

概念对	方法名	相似度	专家值
testis & sperm cell	Resnik	12.345	78.6
	D. Lin	0.5787	
	jcn	jcn	
	λ=0	0.0881	
	λ=0.1	0.0875	
	λ=0.5	0.1133	
	λ=0.7	0.1262	
	λ=1	0.1456	
ovary & oocyte	Resnik	12.345	77.13
	D. Lin	0.4562	
	jcn	0.0409	
	λ=0	0.0977	
	λ=0.1	0.0950	
	λ=0.5	0.1187	
	λ=0.7	0.1306	
	λ=1	0.1485	
organ & cell	Resnik	12.345	84.23
	D. Lin	0.7268	
	jcn	0.0642	
	λ=0	0.1957	
	λ=0.1	0.2109	
	λ=0.5	0.3995	
	λ=0.7	0.4938	
	λ=1	0.6354	

我们做了两组实验。针对只有单一的 ISA 关系的路径，对本文的方法与传统的方法进行比较计算精度；针对存在非 ISA 关系的路径，与传统方法进行适用性比较。RESuLT Project[39] 提供的程序包分别实现了 Resnik，D. Lin，Jiang&Conrath 等方法。本文方法分别设置调节参数 $\lambda=0$，0.1，0.5，0.7，1，并参考了专家给出的评价值，相似度计算结果如表 2 所列。

由表可见，Resnik 方法不能区分具有相同父概念的概念对。在表 2 数据的基础上，参考专家的评价值，我们计算了 pearson 系数，结果如表 3 所列。

表 3 correlation 对比

方法	Resnik	D. Lin	JCn	本文 λ=0	本文 λ=0.1	本文 λ=0.5	本文 λ=0.7	本文 λ=1
相关系数	NaN	0.9633	0.9780	0.9616	0.9685	0.9772	0.9785	0.9796

D. Lin 方法和 JCn 方法与本文方法较接近。在 $\lambda=0$ 时，本文方法等价于语义距离的方法，其相关系数为 0.9616，接近 D. Lin 方法与 JCn 方法；$\lambda=0.1$ 时，本文方法略低于 JCn 方法，但优于 D. Lin 方法；当 $\lambda=0.5$，0.7 时，本文方法的相关系数均大于 D. Lin 方法与 JCn 方法，表明本文提出的方法是较优的；当 $\lambda=1$ 时，本文方法等价于基于概念图信息量覆盖的方法，其相关系数也是最大的，效果最佳。

由表 4 可见，一些传统的方法，如 Resnik，D. lin，JCn 等方法是建立在 ISA 概念关系的基础上的，因此不适用于存在非 ISA 关系的本体。例如图 2 的本体中，概念对〈head，cell〉存在 part-of 关系，根据本文方法，$\lambda=1$ 时，sim(head，cell)=0.5407，本文方法适用。

表 4 存在非 ISA 关系的可用性比较

概念关系	方法名	适用性
非 ISA 关系：Part-of meronym 等	Resnik	不适用
	D. Lin	不适用
	Jiang& Conrath	不适用
	本文方法	适用

结束语 针对传统的本体概念相似度方法不能有效处理非 ISA 概念关系，本文提出了一种新的方法来度量本体中的非 ISA 概念关系。通过概念图信息量覆盖度和语义距离方法，克服了部分传统方法对于一些非 ISA 关系赋予一个特定权值的弊端，本文的方法更具有普遍性和适用性。实验表明，本文方法既可用于 ISA 概念关系，也适用于存在非 ISA 关系的汽车本体和解剖本体。对于一些易出现非 ISA 概念关系的本体，如生物基因本体、制造业本体、农业本体等，我们建议使用本文的方法。

参考文献

[1] Zhong J W，Zhu H P，Li J M，et al. Conceptual graph matching for semantic search[C]// Proceedings of the 10th International Conference on Conceptual Structure. Berlin：Springer-Verlag，2002：92-106

[2] Giunchiglia F，Yatskevich M. Element level semantic matching [C]// Proceedings of Meaning Coordination and Negotiation workshop at ISWC. 2004

[3] Pedersen T，Banerjee S. Maximizing semantic relatedness to perform word sense disambiguation[R]. UMSI 2005/25. 2005

[4] Patwardhan S，Banerjee S，Pedersen T. Using Measures of Se-

mantic Relatedness for Word Sense Disambiguation[C]// Proceedings of 4th International Conference on Computational Linguistics and Intelligent Text Processing. 2003:241-257

[5] Liu M, Shen W M, Hao Q, et al. An weighted ontology-based semantic similarity algorithm for Web service[J]. Expert Systems with Applications, 2009, 36(10): 12480-12490

[6] Tversky A. Features of similarity [J]. Psychological Review, 1977, 84: 327-352

[7] Rada R, Mili H, Bicknell E, et al. Development and application of a metric on semantic nets[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1989, 19(1): 17-30

[8] Wu Z B, Palmer M. Verb semantics and lexical selection[C]// Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics. Las Cruces, New Mexico, 1994: 133-138

[9] Hirst G, David S-O. Lexical chains as representation of context for the detection and correction malapropisms[M]. The MIT Press, 1998

[10] Li Y, Bandar A, Mclean D. An approach for measuring semantic similarity between words using multiple information sources[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4): 871-882

[11] Slimani T, Ben Yaghlane B, Mellouli K. A New Similarity Measure Based on Edge Counting[C]// Proceedings of World Academy of Science, Engineering and Technology. Dec. 2006: 232-236

[12] Cross V. Tversky's Parameterized Similarity Ratio Model: A Basis for Semantic Relatedness[C] // Proceedings of Annual Meeting of the North American Fuzzy Information Processing Society. Montreal Canada, 2006: 541-546

[13] Resnik P. Using information content to evaluate semantic similarity in a taxonomy[C]// Proceedings of the 14th International Joint Conference on Artificial Intelligence. Montreal, 1995: 448-453

[14] Miller G A. WordNet: a lexical database for English[J]. Communications of the ACM(CACM), 1995, 38(11): 39-41

[15] Jiang J, Conrath D. Semantic Similarity based on corpus statistics and lexical taxonomy[C] // Proceedings on International Conference on Research in Computational Linguistics. Taiwan, 1997: 19-33

[16] Lin D. Using syntactic dependency as a local context to resolve word sense ambiguity[C] // Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics. Madrid, 1997: 64-71

[17] Seco N, Veale T, Hayes J. An Intrinsic Information Content Metric for Semantic Similarity in WordNet[C]// Proceedings of 16th European Conference on Artificial Intelligence. Valencia Spain, 2004: 1089-1090

[18] Pirro G. A semantic similarity metric combing features and intrinsic information content[J]. Data & Knowledge Enginnering, 2009, 68(11): 1289-1308

[19] Sanchez D, Batet D, Valls A. Computing Knowledge-based Semantic Similarity from the Web: An Application to the Biomedical Domain[C] // Proceedings of 3rd International Conference on Knowledge Science, Engineering Management. Vienna Austria, vol 5914, 2009: 17-28

[20] Banerjee S, Pedersen T. An adapted Lesk algorithm for word sense disambiguation using WordNet[C] // Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics. Mexico City, 2002

[21] Banerjee S, Pedersen T. Extended gloss overlap as a measure of semantic relatedness[C]// Proceedings of the 18th International Joint Conference on Artificial Intelligence(IJCAI-03). Acapulco, Mexico, 2003: 805-810

[22] Kenneth W C, Patrick H. Word association norms, mutual information, and lexicography [C]// Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics. Vancouver, B. C., 1989: 76-83

[23] Ahmad E S, Hakim H, Abdelkader D Z. Enhancing Semantic Distances With Context Awareness[C] // Proceeding of 8th Journees Francophones Mining and Knowledge Management. Sophia Antipolis, Jan. 2008: 39-49

[24] Danushka B, Yutaka M, Mitsuru I. Measuring Semantic Similarity Between Words Using Web Search Engines[C]// Proceedings of the 16th International Conference on World Wide Web. Banff, Alberta, Canada, 2007: 757-766

[25] Gracia J, Mena E. Web-based Measure of Semantic Relatedness [C]// Proceedings of International Workshop on Web Information Systems Engineering(WISE2008). Auckland New Zealand, 2008: 136-150

[26] Gabrilovich E, Markovitch S. Computing semantic relatedness using wikipedia-based explicit semantic analysis[C]// Proceedings of the Twentieth International Joint Conference for Artificial Intelligence. Hyderabad, India, 2007: 1606-1611

[27] Yin G S, Sheng Q Y. Research on Ontology-based Measuring Semantic Similarity[C] // Proceedings of International Conference on Internet Computing in Science and Engineering. Harbin, China, 2008: 250-253

[28] Michael S. Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network[C] // Proceedings of the Second International Conference on Information and Knowledge Management. Washington, D. C., United States, 1993: 67-74

[29] Xu X H, Huang J L, Wan J, et al. A Method for Measuring Semantic Similarity of Concepts in the Same Ontology[C]// Proceedings of 3rd International Multi-symposiums on Computer and Computational Sciences. Shanghai, China, 2008: 207-213

[30] Shi H X. Research on the Semantic Similarity Computation Method Based on EUO[C]// Proceedings of 2010 Third International Conference on Knowledge Discovery and Data Mining (WKDD2010). 2010: 257-263

[31] Mazuel L, Sabouret N. Semantic relatedness in semantic networks[C]// Proceedings of 18th European Conference on Artificial Intelligence. Univ Patras, Patras Greece, 2008: 727-728

[32] 吴刚，张阔，李涓子，等. 利用相互增强关系迭代计算本体中概念与关系的重要性[J]. 计算机学报，2007，30(9)：1490-1499

[33] Hwang M, Yi H, Choi C, et al. Measurement of Arc-value for Concept similarity[C] // Proceedings of the 7th International Conference on Machine Learning and Cybernetics. Kunming, China, July 2008: 3787-3791

[34] Nguyen H A, Al-Mubaid H. A Combination-based Semantic Similarity Measure Using Multiple Information Sources[C]// Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration. Waikoloa HI, 2006: 617-621

[35] Shu G, Rana O F, Avis N J, et al. Ontology-based semantic matchmaking approach[J]. Advances in Engineering Software, 2007: 38(1): 59-67

[36] 徐德智，吴军庆，陈建二，等. 一种基于概念信息量的相似度传播算法[J]. 计算机科学，2009，36(6)：174-177

[37] Google homepage[EB/OL]. http://www.google.com, 2010-08

[38] Detwiler T, SIG. Foundational Model Explorer [EB/OL]. http://fme.biostr.washington.edu: 8089/FME/index.html, 2003-08-10

[39] Greenwood M A. Pure Java WordNet similarity library, v1.0.0 [CP/OL]. http://nlp.shef.ac.uk.result/software.html, 2007-05-01